PATENT ABSTRACTS OF JAPAN

(11)Publication number : **2000-003187**

(43)Date of publication of application : **07.01.2000**

(51)Int.Cl. **G10L 11/04**
G10K 15/04

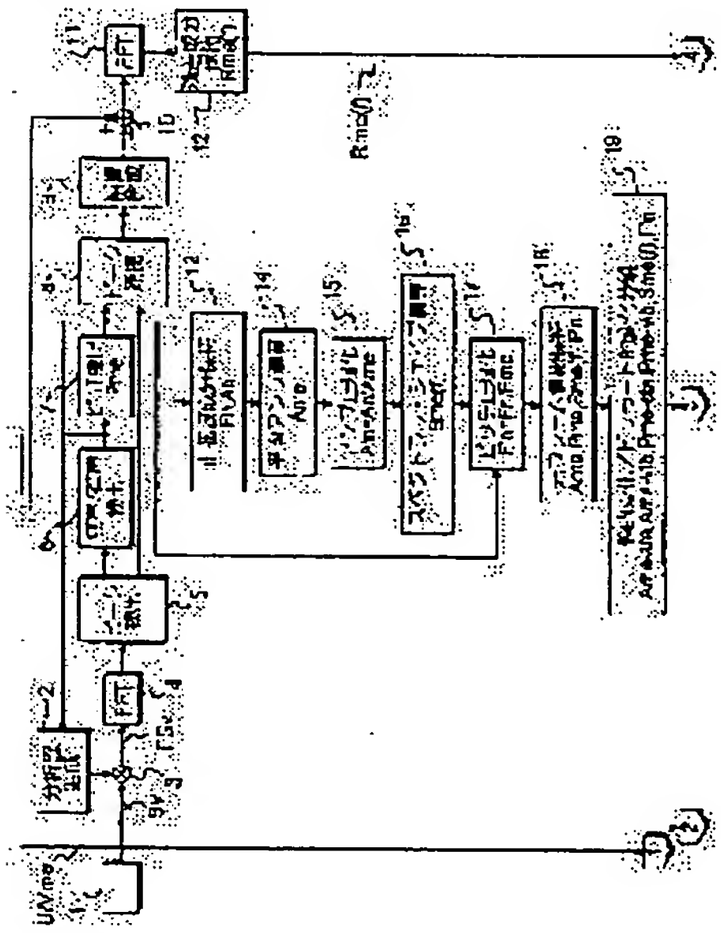
(21)Application number : **10-169046**

(22)Date of filing : **16.06.1998**

(71)Applicant : **YAMAHA CORP**

(72)Inventor : **KONDO TAKAYASU**
XAVIER SERA

(54) METHOD AND DEVICE FOR STORING VOICE FEATURE INFORMATION



(57)Abstract:

PROBLEM TO BE SOLVED: To store precise voice feature information with small storage capacity by obtaining a specified multiplication and a specified difference or ratio related to plural frequencies showing a voice feature.

SOLUTION: An input voice signal segmenting part 3 multiplies an analytic window AW having a period of fixed times of a pitch period generated by an analytic window generation part 2 with an input voice signal Sv. Then, the cut-out part 3 segments the input voice signal Sv in frame to output it to a fast Foulier transform part (FFT) 4 as a frame voice signal FSv. The signal FSv is analytic processed by the FFT 4, and a local peak value shown by combination between a frequency value and an amplifier value is detected from a frequency spectrum being its output by a peak value detection part 5. Plural frequencies Fk (k is natural number) showing such a voice characteristic are obtained, and a reference frequency Fo is multiplied with respective natural numbers (k), and the difference or the ratio between respective multiplication results Fo×k and respective frequencies Fk are obtained to store these difference or the ratio.

Detailed Description of the Invention:

....

[0003]

[Problems to be Solved by the Invention]

In prior-art voice conversion apparatuses, though voice conversion (for example, from male voice to female voice, from female voice to male voice, and the like) is performed, it is simply conversion of voice nature. Therefore, it is impossible to convert voice so that the voice resembles the voice of a particular singer (for example, a professional singer). If there is a function of causing not only voice nature but also the way of singing to resemble those of a particular singer, that is, a function of impersonation, it will be very interesting in a karaoke machine and the like. However, such processing is impossible in prior-art voice conversion apparatuses. Therefore, the inventors provide a voice conversion apparatus capable of causing voice nature to resemble the voice of a target singer.

[0009][2. 4] Step S4

Next, by appropriately selecting and combining attribute data corresponding to a singer (me) trying to do an imitation and target attribute data corresponding to a singer to be imitated, new attribute data (new attribute data = pitch, amplitude and spectrum shape) are obtained.

....

[0010][2. 5] Step S5

Then, based on the subsequently obtained new attribute data, sine wave components of the frame are derived.

[2. 6] Step S6

Then, based on the derived sine wave components and either one of the residual components obtained in step S1 and the residual components of a singer (Target) to be imitated previously memorized (stored), reversed FFT is performed to obtain a conversion voice signals.

[0011][2. 7] Conclusion

By the conversion voice signals resulted from such processing, a reproduced voice becomes just like a singing voice of the other singer (target singer) from a singing voice of a singer trying to do an imitation.

Explanation of Reference numerals:

Figure 1

2 analytic window generation
5 peak detection
6 detection of voiceless/voiced
7 pitch detection
8 peak linkage
9 interpolation/synthesis
12 residual component keeping
13 sine-wave component keeping
14 average amplitude operation
15 amplitude normalization
16 spectral shape operation
17 pitch normalization
18 original frame information keeping
19 static change/vibrato change separation

Figure.2

20 target frame information keeping
21 key control/tempo change
22 easy synchronization processing
23 sine wave component attribute data selection
24 attribute data modification
24' new sine wave component attribute data
25 residual component selection
26 generation of sine wave component
27 sine wave component modification
28 reversed FET
31 sequencer
32 sound generator
34 output

D3

(19)日本国特許庁（ＪＰ）　(12)**公開特許公報**（Ａ）　(11)特許出願公開番号
特開2000－3187
（P2000－3187A）
(43)公開日　平成12年１月７日(2000.1.7)

(51)Int.Cl.'	識別記号	FI	テーマコード'(参考)
Ｇ１０Ｌ　11/04		Ｇ１０Ｌ　9/00	Ｂ　５Ｄ１０８
Ｇ１０Ｋ　15/04	３０２	Ｇ１０Ｋ　15/04	３０２Ｄ

審査請求　未請求　請求項の数５　ＯＬ　（全 15 頁）

(21)出願番号　特願平10－169046

(22)出願日　平成10年６月16日(1998.6.16)

(71)出願人　000004075
ヤマハ株式会社
静岡県浜松市中沢町10番１号
(72)発明者　近藤　高康
静岡県浜松市中沢町10番１号　ヤマハ株式会社内
(72)発明者　ザビエル　セラ
スペイン　バルセロナ　カルデデュー　08440　２－２　ビスカイア19
(74)代理人　100098084
弁理士　川▲崎▼　研二　（外１名）

Ｆターム(参考)　5D108 BF20

(54)【発明の名称】　音声特徴情報記憶方法および音声特徴情報記憶装置

(57)【要約】
【課題】　人間の音声の特徴として周波数成分を記憶する際、メモリ容量を削減する。
【解決手段】　人間の音声信号のを周波数分析すると、複数のローカルピークが観察できる。そして、各ローカルピークの周波数は、ピッチ周波数のほぼ整数倍になる。そこで、ピッチ周波数の整数倍の値と各ローカルピークの周波数との差分または割合を求め、得られた差分または割合を記憶することによりメモリ容量を削減する。

【特許請求の範囲】

【請求項１】　音声の特徴を表わす複数の周波数Ｆｋ（但し、ｋは自然数）を得る過程と、
基準周波数Ｆ0と前記各自然数ｋとの乗算を行う過程と、
各乗算結果Ｆ0×ｋと前記各周波数Ｆｋとの差分または割合を求める過程と、
これら差分または割合を記憶することによって前記音声の特徴を記憶することを特徴とする音声特徴情報記憶方法。

【請求項２】　音声の特徴を表わす複数の周波数Ｆｋ（但し、ｋは自然数）を得る過程と、
基準周波数Ｆ0と前記各自然数ｋとの乗算を行う過程と、
各乗算結果Ｆ0×ｋと前記各周波数Ｆｋとの差分または割合を求める過程と、
これら差分または割合のうち所定のスレッショルド値を超えるものを選択する過程と、
これら選択された差分または割合を記憶することによって前記音声の特徴を記憶することを特徴とする音声特徴情報記憶方法。

【請求項３】　音声の特徴を表わす複数の周波数Ｆｋ（但し、ｋは自然数）を得る過程と、
基準周波数Ｆ0と前記各自然数ｋとの乗算を行う過程と、
各乗算結果Ｆ0×ｋと前記各周波数Ｆｋとの差分または割合を求める過程と、
これら差分または割合のうち大きい順に所定数の差分または割合を選択する過程と、
これら選択された差分または割合を記憶することによって前記音声の特徴を記憶することを特徴とする音声特徴情報記憶方法。

【請求項４】　音声の特徴を表わす複数の周波数Ｆｋ（但し、ｋは自然数）と、これら周波数Ｆｋに対応する振幅値Ａｋを得る過程と、
前記振幅値Ａｋのうち所定値以上であるものに対応する周波数Ｆｋを選択する過程と、
基準周波数Ｆ0と、選択された前記周波数Ｆｋに対応する各自然数ｋとの乗算を行う過程と、
各乗算結果Ｆ0×ｋと前記選択された各周波数Ｆｋとの差分または割合を求める過程と、
これら差分または割合を記憶することによって前記音声の特徴を記憶することを特徴とする音声特徴情報記憶方法。

【請求項５】　請求項１～４の何れかに記載の音声特徴情報記憶方法を実行することを特徴とする音声特徴情報記憶装置。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、音声信号を生成する装置、特にカラオケ装置に用いて好適な音声特徴情報記憶方法および音声特徴情報記憶装置に関する。

【０００２】

【従来の技術】入力された音声の周波数特性などを変えて出力する音声変換装置は種々開発されており、例えば、カラオケ装置の中には、歌い手の歌った歌声のピッチを変換して、男性の声を女性の声に、あるいはその逆に変換させるものもある（例えば、特表平８－５０８５８１号）。

【０００３】

【発明が解決しようとする課題】従来の音声変換装置においては、音声の変換（例えば、男声→女声、女声→男声など）は行われるものの、単に声質を変えるだけに止まっていたので、例えば、特定の歌唱者（例えば、プロの歌手）の声に似せるように変換するということはできなかった。また、声質だけでなく、歌い方までも特定の歌唱者に似させるという、ものまねのような機能があれば、カラオケ装置などにおいては大変に面白いが、従来の音声変換装置ではこのような処理は不可能であった。そこで、本発明者らは、声質を目標（ターゲット）とする歌唱者の声に似させることができる音声変換装置を提供することにした。

【０００４】しかし、かかる装置においては、音声特徴情報を記憶する必要があるため、膨大な記憶容量が必要である。この発明は上述した事情に鑑みてなされたものであり、僅かな記憶容量で高精度な音声特徴情報を記憶できる音声特徴情報記憶方法および音声特徴情報記憶装置を提供することを目的としている。

【０００５】

【課題を解決するための手段】上記課題を解決するため請求項１記載の構成にあっては、音声の特徴を表わす複数の周波数Ｆｋ（但し、ｋは自然数）を得る過程と、基準周波数Ｆ0と前記各自然数ｋとの乗算を行う過程と、各乗算結果Ｆ0×ｋと前記各周波数Ｆｋとの差分または割合を求める過程と、これら差分または割合を記憶することによって前記音声の特徴を記憶することを特徴とする。また、請求項２記載の構成にあっては、音声の特徴を表わす複数の周波数Ｆｋ（但し、ｋは自然数）を得る過程と、基準周波数Ｆ0と前記各自然数ｋとの乗算を行う過程と、各乗算結果Ｆ0×ｋと前記各周波数Ｆｋとの差分または割合を求める過程と、これら差分または割合のうち所定のスレッショルド値を超えるものを選択する過程と、これら選択された差分または割合を記憶することによって前記音声の特徴を記憶することを特徴とする。また、請求項３記載の構成にあっては、音声の特徴を表わす複数の周波数Ｆｋ（但し、ｋは自然数）を得る過程と、基準周波数Ｆ0と前記各自然数ｋとの乗算を行う過程と、各乗算結果Ｆ0×ｋと前記各周波数Ｆｋとの差分または割合を求める過程と、これら差分または割合のうち大きい順に所定数の差分または割合を選択する過

程と、これら選択された差分または割合を記憶することによって前記音声の特徴を記憶することを特徴とする。また、請求項４記載の構成にあっては、音声の特徴を表わす複数の周波数Ｆｋ（但し、ｋは自然数）と、これら周波数Ｆｋに対応する振幅値Ａｋを得る過程と、前記振幅値Ａｋのうち所定値以上であるものに対応する周波数Ｆｋを選択する過程と、基準周波数Ｆ０と、選択された前記周波数Ｆｋに対応する各自然数ｋとの乗算を行う過程と、各乗算結果Ｆ０×ｋと前記選択された各周波数Ｆｋとの差分または割合を求める過程と、これら差分または割合を記憶することによって前記音声の特徴を記憶することを特徴とする。また、請求項５記載の構成にあっては、請求項１～４の何れかに記載の音声特徴情報記憶方法を実行することを特徴とする。

【０００６】

【発明の実施の形態】［１］　実施形態の概要処理

［２］　実施形態の概要処理

次に図面を参照して本発明の好適な実施形態について説明する。始めに、実施形態の概要処理について説明する。

［２．１］　ステップＳ１

まず、ものまねをしようとする歌唱者（me）の音声（入力音声信号）をリアルタイムでＦＦＴ（Fast Fourie Transform）する過程を含むＳＭＳ（Spectral Modeling Synthesis）分析を行い、フレーム単位で正弦波成分（Sine成分）を抽出するとともに、入力音声信号及び正弦波成分からフレーム単位で残差成分（Residual成分）を生成する。これと並行して入力音声信号が無声音（含む無音）か否かを判別し、無声音である場合には、以下のステップＳ２～ステップＳ６の処理は行わず、入力音声信号をそのまま出力することとなる。この場合において、ＳＭＳ分析としては、前回のフレームにおけるピッチに応じて分析窓幅を変更するピッチ同期分析を採用している。

【０００７】［２．２］　ステップＳ２

次に入力音声信号が有声音である場合には、抽出した正弦波成分からさらに元属性（Attribute）データであるピッチ（Pitch）、アンプ（Amplitude）及びスペクトラル・シェイプ（Spectral Shape）を抽出する。さらに抽出したピッチ及びアンプについては、ビブラート成分及びビブラート成分以外の他の成分に分離する。

【０００８】［２．３］　ステップＳ３

予め記憶（保存）してあるものまねの対象（Target）となる歌唱者の属性データ（ターゲット属性データ＝ピッチ、アンプ及びスペクトラル・シェイプ）から、ものまねをしようとする歌唱者（me）の入力音声信号のフレームに対応するフレームのターゲット属性データ（＝ピッチ、アンプ及びスペクトラル・シェイプ）を取り出す。この場合において、ものまねをしようとする歌唱者（me）の入力音声信号のフレームに対応するフレームのターゲット属性データが存在しない場合には、後に詳述するように、予め定めたイージーシンクロナイゼーション規則（Easy Synchronization Rule）に従って、ターゲット属性データを生成し、同様の処理を行う。

【０００９】［２．４］　ステップＳ４

次にものまねをしようとする歌唱者（me）に対応する元属性データ及びものまねの対象となる歌唱者に対応するターゲット属性データを適宜選択して組み合わせることにより、新しい属性データ（新属性データ＝ピッチ、アンプ及びスペクトラル・シェイプ）を得る。なお、ものまねではなく、単なる音声変換として用いる場合には、元属性データ及びターゲット属性データの加算平均として新属性データを得るなどの元属性データ及びターゲット属性データの双方に基づいて計算により新属性データを得るようにすることも可能である。

【００１０】［２．５］　ステップＳ５

つづいて得られた新属性データに基づいて、当該フレームの正弦波成分を求める。

［２．６］　ステップＳ６

そして求めた正弦波成分と、ステップＳ１で求めた残差成分あるいは予め記憶（保存）してあるものまねの対象（Target）となる歌唱者の残差成分のいずれか一方と、に基づいて逆ＦＦＴを行い、変換音声信号を得る。

【００１１】［２．７］　まとめ

これらの処理の結果得られる変換音声信号によれば、再生される音声は、物まねをしようとする歌唱者の歌声が、あたかも、別の歌唱者（ターゲットの歌唱者）が歌った歌声のようになる。

【００１２】［３］　実施形態の詳細構成

図１及び図２に、実施形態の詳細構成図を示す。なお、本実施形態は、本発明による音声変換装置（音声変換方法）をカラオケ装置に適用し、ものまねを行うことができるカラオケ装置として構成した場合の例である。

【００１３】図１において、マイク１は、ものまねをしようとする歌唱者（me）の声を収集し、入力音声信号Ｓｖとして入力音声信号切出部３に出力する。これと並行して、分析窓生成部２は、前回のフレームで検出したピッチの周期の固定倍（例えば、３．５倍など）の周期を有する分析窓（例えば、ハミング窓）ＡＷを生成し、入力音声信号切出部３に出力する。なお、初期状態あるいは前回のフレームが無声音（含む無音）の場合には、予め設定した固定周期の分析窓を分析窓ＡＷとして入力音声信号切出部３に出力する。

【００１４】これらにより入力音声信号切出部３は、入力された分析窓ＡＷと入力音声信号Ｓｖとを掛け合わせ、入力音声信号Ｓｖをフレーム単位で切り出し、フレーム音声信号ＦＳｖとして高速フーリエ変換部４に出力される。

【００１５】より具体的には、入力音声信号Ｓｖとフレームとの関係は、図３に示すようになっており、各フレ

5

ームＦＬは、前のフレームＦＬと一部重なるように設定されている。そして、高速フーリエ変換部４においてフレーム音声信号ＦＳvは、解析処理されるとともに、図４に示すように、高速フーリエ変換部４の出力である周波数スペクトルからピーク検出部５によりローカルピークが検出される。

【００１６】より具体的には、図４に示すような周波数スペクトルに対して、×印を付けたローカルピークを検出する。このローカルピークは、周波数値とアンプ（振幅）値の組み合わせとして表される。すなわち、図４に示すように、（Ｆ０、Ａ０）、（Ｆ１、Ａ１）、（Ｆ２、Ａ２）、……、（ＦＮ、ＡＮ）というように各フレームについてローカルピークが検出され、表されることとなる。

【００１７】そして、図３に模式的に示すように、各フレーム毎に一組（以下、ローカルピーク組という。）として無声／有声検出部６及びピーク連携部８に出力される。無声／有声検出部６は、入力されたフレーム毎のローカルピークに基づいて、高周波成分の大きさに応じて無声であることを検出（‘ｔ’、‘ｋ’等）し、無声／有声検出信号Ｕ／Ｖmeをピッチ検出部７、イージーシンクロナイゼーション処理部２２及びクロスフェーダ部３０に出力する。あるいは、時間軸上で単位時間あたりの零クロス数に応じて無声であることを検出（‘ｓ’等）し、元無声／有声検出信号Ｕ／Ｖmeをピッチ検出部７、イージーシンクロナイゼーション処理部２２及びクロスフェーダ部３０に出力する。

【００１８】さらに無声／有声検出部６は、入力されたフレームについて無声であると検出されなかった場合には、入力されたローカルピーク組をそのまま、ピッチ検出部７に出力する。ピッチ検出部７は、入力されたローカルピーク組に基づいて、当該ローカルピーク組が対応するフレームのピッチＰmeを検出する。

【００１９】より具体的なフレームのピッチＰmeの検出方法としては、例えば、Maher,R.C.andJ.W.Beauchamp:"Fundamental Frequency Estimation of Musical Signal using a two-way Mismatch Procedure"（Journal of Acounstical Society of America95(4):2254-2263）に開示されているような方法で行う。

【００２０】次に、ピーク検出部５から出力されたローカルピーク組は、ピーク連携部８において、前後のフレームについて連携が判断され、連携すると認められるローカルピークについては、一連のデータ列となるようにローカルピークをつなげる連携処理がなされる。

【００２１】ここで、この連携処理について、図５を参照して説明する。今、図５（Ａ）に示すようなローカルピークが前回のフレームにおいて検出され、図５（Ｂ）に示すようなローカルピークが今回のフレームにおいて検出されたとする。

【００２２】この場合、ピーク連携部８は、前回のフレームで検出された各ローカルピーク（Ｆ０、Ａ０）、（Ｆ１、Ａ１）、（Ｆ２、Ａ２）、……、（ＦＮ、ＡＮ）に対応するローカルピークが今回のフレームでも検出されたか否かを調べる。対応するローカルピークがあるか否かの判断は、前回のフレームで検出されたローカルピークの周波数を中心にした所定範囲内に今回のフレームのローカルピークが検出されるか否かによって行われる。

【００２３】より具体的には、図５の例では、ローカルピーク（Ｆ０、Ａ０）、（Ｆ１、Ａ１）、（Ｆ２、Ａ２）……については、対応するローカルピークが検出されているが、ローカルピーク（ＦＫ、ＡＫ）については（図５（Ａ）参照）、対応するローカルピーク（図５（Ｂ）参照）は検出されていない。

【００２４】ピーク連携部８は、対応するローカルピークを検出した場合は、それらを時系列順に繋げて一組のデータ列として出力する。なお、対応するローカルピークが検出されない場合は、当該フレームについての対応ローカルピークは無しということを示すデータに置き換える。ここで、図６は、複数のフレームにわたるローカルピークの周波数Ｆ０及び周波数Ｆ１の変化の一例を示している。

【００２５】このような変化は、アンプ（振幅）Ａ０、Ａ１、Ａ２、……についても同様に認められる。この場合、ピーク連携部８から出力されるデータ列は、フレームの間隔おきに出力される離散的な値である。なお、ピーク連携部８から出力されるピーク値を、以後において、確定成分という。これは、元の信号（すなわち、音声信号Ｓｖ）のうち正弦波の要素として確定的に置き換えられる成分という意味である。また、置き換えられた各正弦波（厳密には、正弦波のパラメータである周波数及びアンプ（振幅））の各々については、正弦波成分と呼ぶことにする。

【００２６】次に、補間合成部９は、ピーク連携部８から出力される確定成分について補間処理を行い、補間後の確定成分に基づいていわゆるオシレータ方式で波形合成を行う。この場合の補間の間隔は、後述する出力部３４が出力する最終出力信号のサンプリングレート（例えば、４４．１ＫＨｚ）に対応した間隔で行われる。前述した図６に示す実線は、正弦波成分の周波数Ｆ０、Ｆ１について補間処理が行われた場合のイメージを示している。

【００２７】［３．１］　補間合成部の構成

ここで、補間合成部９の構成を図７に示す。補間合成部９は、複数の部分波形発生部９ａを備えて構成されており、各部分波形発生部９ａは、指定された正弦波成分の周波数（Ｆ０、Ｆ１、…）およびアンプ（振幅）に応じた正弦波を発生する。ただし、本第１実施形態における正弦波成分（Ｆ０、Ａ０）、（Ｆ１、Ａ１）、（Ｆ２、Ａ２）、……は、各々補間の間隔に従って時事刻々変化

6

していくものであるから、各部分波形発生部９ａから出力される波形は、その変化に従った波形になる。

【００２８】すなわち、ピーク連携部８からは正弦波成分（Ｆ０、Ａ０）、（Ｆ１、Ａ１）、（Ｆ２、Ａ２）、……が順次出力され、各正弦波成分の各々について補間処理が行われるから、各部分波形発生部９ａは、所定の周波数領域内で周波数と振幅が変動する波形を出力する。そして、各部分波形発生部９ａから出力された波形は、加算部９ｂにおいて加算合成される。したがって、補間合成部９の出力信号は、入力音声信号Ｓｖから確定成分を抽出した正弦波成分合成信号ＳＳＳになる。

【００２９】［３．１．１］　補間合成部９のデータ構成

ここで、補間合成部９におけるデータ構成について説明する。補間合成部９は各部分成分について周波数とアンプのペアをＮ＋１組有しており、高い精度でこれらのデータを記憶しようとすると、膨大なメモリ容量が必要になる。一方、データの有効桁を少なくする等の手法により精度を下げると、音声信号の忠実度も下がる。

【００３０】一方、人間の音声信号の性質として、各部分成分の周波数は、ピッチ周波数のほぼ整数倍になる。そこで、この性質を利用して、ピッチ周波数の整数倍の値と各部分成分の周波数との相違に着目すれば、少ないメモリ容量で忠実な再生が可能であると考えられる。具体的には、以下に述べる何れかの方法、またはこれらの組み合わせを採用すると好適である。

【００３１】（１）差分を記憶する方法

周波数Ｆｋ（但しk=0～N）は、以下のように表わすことができる。

Ｆｋ＝Ｆ０×ｋ＋ｄＦｋ

ここで、Ｆ０×ｋは、ピッチ周波数の整数倍の値であり、ｄＦｋは、この整数倍の値と実際の周波数Ｆｋとの差分値である。実際に周波数Ｆｋを記憶せず、差分値ｄＦｋを記憶しておくことにより、周波数Ｆｋを記憶するためのメモリ容量を削減することができる。但し、差分値ｄＦｋの取りうる可能性のある値の最小値および最大値は、値「ｋ」に比例して増加する。

【００３２】（２）比率を記憶する方法

周波数Ｆｋ（但しk=0～N）は、以下の式によっても表わすことができる。

Ｆｋ－Ｆ０×ｋ×ｒＦｋ

ここで、ｒＦｋは、ピッチ周波数の整数倍の値Ｆ０×ｋと実際の周波数Ｆｋとの比率である。そこで、実際に周波数Ｆｋを記憶せず、比率ｒＦｋを記憶しておくことにより、周波数Ｆｋを記憶するためのメモリ容量を削減することができる。この比率ｒＦｋは、値「ｋ」に関係なくほぼ一定である点で上述した差分値ｄＦｋよりも扱いが容易である。

【００３３】ここで、上記各周波数Ｆｋの採りうる範囲は５０Ｈｚ～１０ｋＨｚ程度確保しておけば充分である。一方、比率ｒＦｋの範囲は個人差があるが、本発明者らが観測したところによれば、１音程（１００セント）程度確保しておけば大部分の人の音声を忠実に再現することができる。また、比率ｒＦｋの精度は１セント程度確保しておけば充分である。

【００３４】（３）比率を対数値で記憶する方法

上記比率ｒＦｋを記憶する際、これを対数値に変換しておくと、メモリ容量を一層削減することができる。この対数値ｃＦｋとして「セント」を用いるとすれば、対数値ｃＦｋは下式により求まる。

ｃＦｋ＝１２００×log2(ｒＦｋ)

具体的には、対数値ｃＦｋによって＋／－１００セントの範囲を１セントの精度で表現するためには、対数値ｃＦｋを８ビットで記憶させるとよい。

【００３５】（４）一部の周波数Ｆｋの記憶を省略する方法

この方法は、上記方法(１)～(３)の何れかの方法と組み合わせて採られうる方法である。上記方法(１)～(３)においては、「（Ｎ＋１）個」の差分値ｄＦｋ、比率ｒＦｋまたは対数値ｃＦｋが必要であると考えられる。しかし、本発明者らの実験によれば、一部の周波数Ｆｋについてはピッチ周波数Ｆ０の整数倍であると仮定したとしても音質上の劣化が少ないことが判明した。かかる部分成分においては、周波数Ｆｋとしてピッチ周波数Ｆ０の整数倍の値を用いることができ、対数値ｃＦｋ等を記憶する必要が無くなる。

【００３６】方法(３)，(４)を組み合わせて採用することを想定すると、周波数Ｆｋを忠実に再現すべき部分成分は、以下の方法(４．１)～（４．３）を採用して決定することができる。なお、方法(４．１)～（４．３）は単独で用いてもよく、組み合わせで用いてもよい。

(４．１)再現数Ｍ（但しＭ＜Ｎ＋１）を予め決定しておき、対数値ｃＦｋの大きい順にＭ個の部分成分を選択する。

(４．２)対数値ｃＦｋに対してスレッショルド値を決定しておき、対数値ｃＦｋが該スレッショルド値を超えた部分成分を選択する。

(４．３)アンプの大きさが所定の条件（例えば最大のＡｋに対して－３０ｄＢよりも大きい値）を満たす部分成分を選択する。

【００３７】以上のように選択された部分成分に係る周波数情報をＭ個記憶する場合、メモリの所定の領域に値Ｍを記憶し、何番目の成分に対応するかを示す値ｋと、周波数を特定するための情報（上記対数値ｃＦｋ等）とをＭ組記憶するとよい。かかる方法は、部分成分数Ｎ＋１よりも再現数Ｍがかなり小さい場合に特に有効である。

【００３８】（５）アンプＡｋの記憶方法

この方法は、上記方法(４)またはこれと方法(１)～(３)とを組み合わせて採られうる方法である。本実施形態においては、上述したように各部分成分に対してアンプＡ

kが記憶される。各アンプＡkに対して１バイト（８ビット）を割り当て、データの精度を１ｄＢにすると、アンプＡkは２５６ｄＢのダイナミックレンジを有することになる。しかし、実際にはこのように広いダイナミックレンジは不要であり、１２８ｄＢ程度確保できれば充分である。

【００３９】通常のコンピュータにおいては、データ長を８ビット単位で設定するため、アンプＡkに７ビットを割り当てて１２８ｄＢのダイナミックレンジを確保すると、１ビット余剰が生じることになる。そこで、この１ビットにおいて、周波数を特定するための情報（上記対数値cＦｋ等）が存在するか否かを示すことにする。以下、この情報をフラグｘｋという

【００４０】そうすると、何番目の成分に対応するかを示す値ｋを記憶するために独立の記憶領域を設ける必要が無くなるため、メモリ容量を一層削減することができる。なお、アンプＡkのデシベル値の表現の仕方についても種々の態様が考えられる。例えば、アンプＡkのうち最大値を０ｄＢとして、０～－１２７ｄＢの範囲で表現してもよい。また、最大値が０ｄＢを超えた値を持つ場合、あるいは、必要なダイナミックレンジが狭く高い分解能が望まれる場合は、下式によりアンプｎＡkを求め、アンプＡkに代えてアンプｎＡkを記憶してもよい。

ｎＡk＝α・（Ａk＋β）

【００４１】すなわち、上式においてアンプｎＡkが０～１２７の範囲に収まるようにα又はβを決定するとよい。例として必要なダイナミックレンジが＋２０ｄＢ～－４０ｄＢであった場合、α＝－１２７／６０、β＝－２０とすると、アンプＡkが２０ｄＢの時はｎＡｋ＝０、アンプＡkが－１０ｄＢの時はｎＡｋ＝１２７となり、記憶エリアを有効に利用できる。αおよびβの値は予め決定しておき固定値にしてもよく、状況に応じて変化させたい場合は一方、または双方の値もデータ列に加える等の措置により、可変にしてもよい。但し、上記の方法において０～１２７の範囲を超えたデータが存在した場合は、その範囲に入るように、０以下の値は０に、１２７以上の値は１２７に揃えることは必要であろう。

【００４２】（６）まとめ

上記方法(３)、(４)および(５)を総合すると、周波数Ｆkを忠実に再現すべき正弦波成分においては、アンプＡｋ（７ビット）と、フラグｘｋ（１ビット）と、対数値cＦｋ（８ビット）とによって合計１６ビットのメモリ容量が必要になる。また、周波数Ｆｋをピッチ周波数Ｆ０の整数倍に近似して良い場合は、アンプＡｋ（７ビット）と、フラグｘｋ（１ビット）とによって合計８ビットのメモリ容量が必要になる。一般的な浮動小数点データは３２ビット長が必要であるから、アンプＡkおよび周波数Ｆkを合わせると、正弦波成分あたり６４ビットが必要になる。本実施形態においては、これを８～１６ビットに削減できるから、所要メモリ容量は１／８～１／４程度に削減することが可能である。

【００４３】［３．２］　残差成分検出部の動作

次に、残差成分検出部１０は、補間合成部９から出力された正弦波成分合成信号ＳSSと入力音声信号Ｓｖとの偏差である残差成分信号ＳRD（時間波形）を生成する。この残差成分信号ＳRDは、音声に含まれる無声成分を多く含む。一方、前述の正弦波成分合成信号ＳSSは有声成分に対応するものである。ところで、目標（Target）となる歌唱者の声に似せるには、有声音についてだけ処理を行えば、無声音については処理を施す必要はあまりない。そこで、本実施形態においては、有声母音成分に対応する確定成分について音声変換処理を行うようにしている。より具体的には、残差成分信号ＳRDについては、高速フーリエ変換部１１で、周波数波形に変換し、得られた残差成分信号（周波数波形）をＲme(f)として残差成分保持部１２に保持しておく。

【００４４】［３．３］　平均アンプ演算部の動作

一方、図８（Ａ）に示すように、ピーク検出部５からピーク連携部８を介して出力された正弦波成分（Ｆ０、Ａ０）、（Ｆ１、Ａ１）、（Ｆ２、Ａ２）、……、（Ｆ(N-1)、Ａ(N-1)）のＮ個の正弦波成分（以下、これらをまとめてＦｎ、Ａｎと表記する。ｎ＝０～（Ｎ－１）。）は、正弦波成分保持部１３に保持されるとともに、アンプＡｎは平均アンプ演算部１４に入力され、各フレーム毎に次式により平均アンプＡmeが算出される。

Ａme＝Σ（Ａｎ）／Ｎ

【００４５】［３．４］　アンプ正規化部の動作

次にアンプ正規化部１５において、次式により各アンプＡｎを平均アンプＡmeで正規化し、正規化アンプＡ’ｎを求める。

Ａ’ｎ＝Ａｎ／Ａme

【００４６】［３．５］　スペクトラル・シェイプ演算部の動作

そして、スペクトラル・シェイプ演算部１６において、図８（Ｂ）に示すように、周波数Ｆｎ及び正規化アンプＡ’ｎにより得られる正弦波成分（Ｆｎ、Ａ’ｎ）をブレークポイントとするエンベロープ（包絡線）をスペクトラル・シェイプＳme(f)として生成する。この場合において、二つのブレークポイント間の周波数におけるアンプの値は、当該二つのブレークポイントを、例えば、直線補間することにより算出する。なお、補間の方法は直線補間に限られるものではない。

【００４７】［３．６］　ピッチ正規化部の動作

続いてピッチ正規化部１７においては、各周波数Ｆｎをピッチ検出部７において検出したピッチＰmeで正規化し、正規化周波数Ｆ’ｎを求める。Ｆ’ｎ＝Ｆｎ／Ｐme

これらの結果、元フレーム情報保持部１８は、入力音声信号Ｓvに含まれる正弦波成分に対応する元属性データである平均アンプＡme、ピッチＰme、スペクトラル・シェイプＳme(f)、正規化周波数Ｆ’ｎを保持することと

なる。

【００４８】なお、この場合において、正規化周波数Ｆ’ｎは、倍音列の周波数の相対値を表しており、もし、フレームの倍音構造を完全倍音構造であるとして取り扱うならば、保持する必要はない。この場合において、男声／女声変換を行おうとしている場合には、この段階において、男声→女声変換を行う場合には、ピッチをオクターブ上げ、女声→男声変換を行う場合にはピッチをオクターブ下げる男声／女声ピッチ制御処理を行うようにするのが好ましい。

【００４９】つづいて、元フレーム情報保持部１８に保持している元属性データのうち、平均アンプＡmeおよびピッチＰmeについては、さらに静的変化／ビブラート的変化分離部１９により、フィルタリング処理などを行って、静的変化成分とビブラート変化的成分とに分離して保持する。なお、さらにビブラート変化的成分からより高周波変化成分であるジッタ変化的成分を分離するように構成することも可能である。

【００５０】より具体的には、平均アンプＡmeを平均アンプ静的成分Ａme-sta及び平均アンプビブラート的成分Ａme-vibとに分離して保持する。また、ピッチＰmeをピッチ静的成分Ｐme-sta及びピッチビブラート的成分Ｐme-vibとに分離して保持する。

【００５１】これらの結果、対応するフレームの元フレーム情報データＩＮＦmeは、図８（Ｃ）に示すように、入力音声信号Ｓvの正弦波成分に対応する元属性データである平均アンプ静的成分Ａme-sta、平均アンプビブラート的成分Ａme-vib、ピッチ静的成分Ｐme-sta、ピッチビブラート的成分Ｐme-vib、スペクトラル・シェイプＳme(f)、正規化周波数Ｆ’ｎ及び残差成分Ｒme（ｆ）の形で保持されることとなる。

【００５２】一方、ものまねの対象（target）となる歌唱者に対応するターゲット属性データから構成されるターゲットフレーム情報データＩＮＦtarは、予め分析されてターゲットフレーム情報保持部２０を構成するハードディスクなどに予め保持されている。この場合において、ターゲットフレーム情報データＩＮＦtarのうち、正弦波成分に対応するターゲット属性データとしては、平均アンプ静的成分Ａtar-sta、平均アンプビブラート的成分Ａtar-vib、ピッチ静的成分Ｐtar-sta、ピッチビブラート的成分Ｐtar-vib、スペクトラル・シェイプＳtar(f)がある。

【００５３】また、ターゲットフレーム情報データＩＮＦtarのうち、残差成分に対応するターゲット属性データとしては、残差成分Ｒtar(f)がある。

【００５４】［３．７］　キーコントロール／テンポチェンジ部の動作

次にキーコントロール／テンポチェンジ部２１は、シーケンサ３１からの同期信号ＳSYNCに基づいて、ターゲットフレーム情報保持部２０から同期信号ＳSYNCに対応するフレームのターゲットフレーム情報ＩＮＦtarの読出処理及び読み出したターゲットフレーム情報データＩＮＦtarを構成するターゲット属性データの補正処理を行うとともに、読み出したターゲットフレーム情報ＩＮＦtarおよび当該フレームが無声であるか有声であるかを表すターゲット無声／有声検出信号Ｕ／Ｖtarを出力する。

【００５５】より具体的には、キーコントロール／テンポチェンジ部２１の図示しないキーコントロールユニットは、カラオケ装置のキーを基準より上げ下げした場合、ターゲット属性データであるピッチ静的成分Ｐtar-sta及びピッチビブラート的成分Ｐtar-vibについても、同じだけ上げ下げする補正処理を行う。例えば、５０［cent］だけキーを上げた場合には、ピッチ静的成分Ｐtar-sta及びピッチビブラート的成分Ｐtar-vibについても５０［cent］だけ上げなければならない。

【００５６】また、キーコントロール／テンポチェンジ部２１の図示しないテンポチェンジユニットは、カラオケ装置のテンポを上げ下げした場合には、変更後のテンポに相当するタイミングで、ターゲットフレーム情報データＩＮＦtarの読み出し処理を行う必要がある。この場合において、必要なフレームに対応するタイミングに相当するターゲットフレーム情報データＩＮＦtarが存在しない場合には、当該必要なフレームのタイミングの前後のタイミングに存在する二つのフレームのターゲットフレーム情報データＩＮＦtarを読み出し、これら二つのターゲットフレーム情報データＩＮＦtarにより補間処理を行い、当該必要なタイミングにおけるフレームのターゲットフレーム情報データＩＮＦtar、ひいては、ターゲット属性データを生成する。

【００５７】この場合において、ビブラート的成分（平均アンプビブラート的成分Ａtar-vib及びピッチビブラート的成分Ｐtar-vib）に関しては、そのままでは、ビブラートの周期自体が変化してしまい、不適当であるので、周期が変動しないような補間処理を行う必要がある。又は、ターゲット属性データとして、ビブラートの軌跡そのものを表すデータではなく、ビブラート周期及びビブラート深さのパラメータを保持し、実際の軌跡を演算により求めるようにすれば、この不具合を回避することができる。

【００５８】［３．８］　イージーシンクロナイゼーション処理部の動作

次にイージーシンクロナイゼーション処理部２２は、ものまねをしようとする歌唱者のフレーム（以下、元フレームという。）に元フレーム情報データＩＮＦmeが存在するにもかかわらず、対応するものまねの対象となる歌唱者のフレーム（以下、ターゲットフレームという。）にターゲットフレーム情報データＩＮＦtarが存在しない場合には、当該ターゲットフレームの前後方向に存在するフレームのターゲットフレーム情報データＩＮＦta

rを当該ターゲットフレームのターゲットフレーム情報データＩＮＦtarとするイージーシンクロナイゼーション処理を行う。

【００５９】そして、イージーシンクロナイゼーション処理部２２は、後述する置換済ターゲットフレーム情報データＩＮＦtar-syncに含まれるターゲット属性データのうち正弦波成分に関するターゲット属性データ（平均アンプ静的成分Ａtar-sync-sta、平均アンプビブラート的成分Ａtar-sync-vib、ピッチ静的成分Ｐtar-sync-sta、ピッチビブラート的成分Ｐtar-sync-vib及びスペクトラル・シェイプＳtar-sync(f)）を正弦波成分属性データ選択部２３に出力する。

【００６０】また、イージーシンクロナイゼーション処理部２２は、後述する置換済ターゲットフレーム情報データＩＮＦtar-syncに含まれるターゲット属性データのうち残差成分に関するターゲット属性データ（残差成分Ｒtar-sync(f)）を残差成分選択部２５に出力する。

【００６１】このイージーシンクロナイゼーション処理部２２における処理においても、ビブラート的成分（平均アンプビブラート的成分Ａtar-vib及びピッチビブラート的成分Ｐtar-vib）に関しては、そのままでは、ビブラートの周期自体が変化してしまい、不適当であるので、周期が変動しないような補間処理を行う必要がある。又は、ターゲット属性データとして、ビブラートの軌跡そのものを表すデータではなく、ビブラート周期及びビブラート深さのパラメータを保持し、実際の軌跡を演算により求めるようにすれば、この不具合を回避することができる。

【００６２】［３．８．１］　イージーシンクロナイゼーション処理の詳細

ここで、図９及び図１０を参照してイージーシンクロナイゼーション処理について詳細に説明する。図９は、イージーシンクロナイゼーション処理のタイミングチャートであり、図１０はイージーシンクロナイゼーション処理フローチャートである。

【００６３】まず、イージーシンクロナイゼーション処理部２２は、シンクロナイゼーション処理の方法を表すシンクロナイゼーションモード＝“０”とする（ステップＳ１１）。このシンクロナイゼーションモード＝“０”は、元フレームに対応するターゲットフレームにターゲットフレーム情報データＩＮＦtarが存在する通常処理の場合に相当する。

【００６４】そしてあるタイミングｔにおける元無声／有声検出信号Ｕ／Ｖme(t)が無声（Ｕ）から有声（Ｖ）に変化したか否かを判別する（ステップＳ１２）。例えば、図９に示すように、タイミングｔ＝ｔ1においては、元無声／有声検出信号Ｕ／Ｖme(t)が無声（Ｕ）から有声（Ｖ）に変化している。

【００６５】ステップＳ１２の判別において、元無声／有声検出信号Ｕ／Ｖme(t)が無声（Ｕ）から有声（Ｖ）に変化している場合には（ステップＳ１２；Ｙｅｓ）、タイミングｔの前回のタイミングｔ-1における元無声／有声検出信号Ｕ／Ｖme(t-1)が無声（Ｕ）かつターゲット無声／有声検出信号Ｕ／Ｖtar(t-1)が無声（Ｕ）であるか否かを判別する（ステップＳ１８）。例えば、図９に示すように、タイミングｔ＝ｔ０（＝ｔ1-1）においては、元無声／有声検出信号Ｕ／Ｖme(t-1)が無声（Ｕ）かつターゲット無声／有声検出信号Ｕ／Ｖtar(t-1)が無声（Ｕ）となっている。

【００６６】ステップＳ１８の判別において、元無声／有声検出信号Ｕ／Ｖme(t-1)が無声（Ｕ）かつターゲット無声／有声検出信号Ｕ／Ｖtar(t-1)が無声（Ｕ）となっている場合には（ステップＳ１８；Ｙｅｓ）、当該ターゲットフレームには、ターゲットフレーム情報データＩＮＦtarが存在しないので、シンクロナイゼーションモード＝“１”とし、置換用のターゲットフレーム情報データＩＮＦholdを当該ターゲットフレームの後方向（Backward）に存在するフレームのターゲットフレーム情報とする。

【００６７】例えば、図９に示すように、タイミングｔ＝ｔ1～ｔ2のターゲットフレームには、ターゲットフレーム情報データＩＮＦtarが存在しないので、シンクロナイゼーションモード＝“１”とし、置換用ターゲットフレーム情報データＩＮＦholdを当該ターゲットフレームの後方向に存在するフレーム（すなわち、タイミングｔ＝ｔ2～ｔ3に存在するフレーム）のターゲットフレーム情報データbackwardとする。

【００６８】そして、処理をステップＳ１５に移行し、シンクロナイゼーションモード＝“０”であるか否かを判別する（ステップＳ１５）。ステップＳ１５の判別において、シンクロナイゼーションモード＝“０”である場合には、タイミングｔにおける元フレームに対応するターゲットフレームにターゲットフレーム情報データＩＮＦtar(t)が存在する場合、すなわち、通常処理であるので、置換済ターゲットフレーム情報データＩＮＦtar-syncをターゲットフレーム情報データＩＮＦtar(t)とする。

【００６９】ＩＮＦtar-sync＝ＩＮＦtar(t)

例えば、図９に示すようにタイミングｔ＝ｔ2～ｔ3のターゲットフレームには、ターゲットフレーム情報データＩＮＦtarが存在するので、

ＩＮＦtar-sync＝ＩＮＦtar(t)

とする。

【００７０】この場合において、以降の処理に用いられる置換済ターゲットフレーム情報データＩＮＦtar-syncに含まれるターゲット属性データ（平均アンプ静的成分Ａtar-sync-sta、平均アンプビブラート的成分Ａtar-sync-vib、ピッチ静的成分Ｐtar-sync-sta、ピッチビブラート的成分Ｐtar-sync-vib、スペクトラル・シェイプＳtar-sync(f)及び残差成分Ｒtar-sync(f)）は実質的に

は、以下の内容となる（ステップＳ１６）。

【００７１】

A tar-sync-sta= A tar-sta

A tar-sync-vib= A tar-vib

P tar-sync-sta= P tar-sta

P tar-sync-vib= P tar-vib

S tar-sync(f)= S tar(f)

R tar-sync(f)= R tar(f)

【００７２】ステップＳ１５の判別において、シンクロナイゼーションモード＝“１”またはシンクロナイゼーションモード＝“２”である場合には、タイミングｔにおける元フレームに対応するターゲットフレームにターゲットフレーム情報データＩＮＦtar(t)が存在しない場合であるので、置換済ターゲットフレーム情報データＩＮＦtar-syncを置換用ターゲットフレーム情報データＩＮＦholdとする。

【００７３】ＩＮＦtar-sync＝ＩＮＦhold

例えば、図９に示すように、タイミングｔ＝ｔ1～ｔ2のターゲットフレームには、ターゲットフレーム情報データＩＮＦtarが存在せず、シンクロナイゼーションモード＝“１”となるが、タイミングｔ＝ｔ2～ｔ3のターゲットフレームには、ターゲットフレーム情報データＩＮＦtarが存在するので、置換済ターゲットフレーム情報データＩＮＦtar-syncをタイミングｔ＝ｔ2～ｔ3のターゲットフレームのターゲットフレーム情報データである置換用ターゲットフレーム情報データＩＮＦholdとする処理Ｐ１を行い、以降の処理に用いられる置換済ターゲットフレーム情報データＩＮＦtar-syncに含まれるターゲット属性データは、平均アンプ静的成分A tar-sync-sta、平均アンプビブラート的成分A tar sync vib、ピッチ静的成分P tar-sync-sta、ピッチビブラート的成分P tar-sync-vib、スペクトラル・シェイプS tar-sync(f)及び残差成分R tar-sync(f)となる（ステップＳ１６）。

【００７４】また、図９に示すように、タイミングｔ＝ｔ3～ｔ4のターゲットフレームには、ターゲットフレーム情報データＩＮＦtarが存在せず、シンクロナイゼーションモード＝“２”となるが、タイミングｔ＝ｔ2～ｔ3のターゲットフレームには、ターゲットフレーム情報データＩＮＦtarが存在するので、置換済ターゲットフレーム情報データＩＮＦtar-syncをタイミングｔ－ｔ2～ｔ3のターゲットフレームのターゲットフレーム情報データである置換用ターゲットフレーム情報データＩＮＦholdとする処理Ｐ２を行い、以降の処理に用いられる置換済ターゲットフレーム情報データＩＮＦtar-syncに含まれるターゲット属性データは、平均アンプ静的成分A tar-sync-sta、平均アンプビブラート的成分A tar-sync-vib、ピッチ静的成分P tar-sync-sta、ピッチビブラート的成分P tar-sync-vib、スペクトラル・シェイプS tar-sync(f)及び残差成分R tar-sync(f)となる（ステップＳ１６）。

【００７５】ステップＳ１２の判別において、元無声／有声検出信号Ｕ／Ｖme(t)が無声（Ｕ）から有声（Ｖ）に変化していない場合には（ステップＳ１２；Ｎｏ）、ターゲット無声／有声検出信号Ｕ／Ｖtar(t)が有声（Ｖ）から無声（Ｕ）に変化しているか否かを判別する（ステップＳ１３）。ステップＳ１３の判別において、ターゲット無声／有声検出信号Ｕ／Ｖtar(t)が有声（Ｖ）から無声（Ｕ）に変化している場合には（ステップＳ１３；Ｙｅｓ）、タイミングｔの前回のタイミングｔ-1における元無声／有声検出信号Ｕ／Ｖme(t-1)が有声（Ｖ）かつターゲット無声／有声検出信号Ｕ／Ｖtar(t-1)が有声（Ｖ）であるか否かを判別する（ステップＳ１９）。

【００７６】例えば、図９に示すように、タイミングｔ3においてターゲット無声／有声検出信号Ｕ／Ｖtar(t)が有声（Ｖ）から無声（Ｕ）に変化し、タイミングｔ-1＝ｔ2～ｔ3においては、元無声／有声検出信号Ｕ／Ｖme(t-1)が有声（Ｖ）かつターゲット無声／有声検出信号Ｕ／Ｖtar(t-1)が有声（Ｕ）となっている。ステップＳ１８の判別において、元無声／有声検出信号Ｕ／Ｖme(t-1)が有声（Ｖ）かつターゲット無声／有声検出信号Ｕ／Ｖtar(t-1)が有声（Ｖ）となっている場合には（ステップＳ１９；Ｙｅｓ）、当該ターゲットフレームには、ターゲットフレーム情報データＩＮＦtarが存在しないので、シンクロナイゼーションモード＝“２”とし、置換用のターゲットフレーム情報データＩＮＦholdを当該ターゲットフレームの前方向（forward）に存在するフレームのターゲットフレーム情報とする。

【００７７】例えば、図９に示すように、タイミングｔ＝ｔ3～ｔ4のターゲットフレームには、ターゲットフレーム情報データＩＮＦtarが存在しないので、シンクロナイゼーションモード＝“２”とし、置換用ターゲットフレーム情報データＩＮＦholdを当該ターゲットフレームの前方向に存在するフレーム（すなわち、タイミングｔ＝ｔ2～ｔ3に存在するフレーム）のターゲットフレーム情報データforwardとする。

【００７８】そして、処理をステップＳ１５に移行し、シンクロナイゼーションモード＝“０”であるか否かを判別して（ステップＳ１５）、以下、同様の処理を行う。ステップＳ１３の判別において、ターゲット無声／有声検出信号Ｕ／Ｖtar(t)が有声（Ｖ）から無声（Ｕ）に変化していない場合には（ステップＳ１３；Ｎｏ）、タイミングｔにおける元無声／有声検出信号Ｕ／Ｖme(t)が有声（Ｖ）から無声（Ｕ）に変化し、あるいは、ターゲット無声／有声検出信号Ｕ／Ｖtar(t)が無声（Ｕ）から有声（Ｖ）に変化しているか否かを判別する（ステップＳ１４）。

【００７９】ステップＳ１４の判別において、タイミングｔにおける元無声／有声検出信号Ｕ／Ｖme(t)が有声

（Ｖ）から無声（Ｕ）に変化し、かつ、ターゲット無声／有声検出信号Ｕ／Ｖtar(t)が無声（Ｕ）から有声（Ｖ）に変化している場合には（ステップＳ１４；Ｙｅｓ）、シンクロナイゼーションモード＝“０”とし、置換用ターゲットフレーム情報データＩＮＦholdを初期化（clear）し、処理をステップＳ１５に移行して、以下、同様の処理を行う。ステップＳ１４の判別において、タイミングｔにおける元無声／有声検出信号Ｕ／Ｖme(t)が有声（Ｖ）から無声（Ｕ）に変化せず、あるいは、ターゲット無声／有声検出信号Ｕ／Ｖtar(t)が無声（Ｕ）から有声（Ｖ）に変化していない場合には（ステップＳ１４；Ｎｏ）、そのまま処理をステップＳ１５に移行し、以下同様の処理を行う。

【００８０】［３．９］　正弦波成分属性データ選択部の動作

続いて、正弦波成分属性データ選択部２３は、イージーシンクロナイゼーション処理部２２から入力された置換済ターゲットフレーム情報データＩＮＦtar-syncに含まれるターゲット属性データのうち正弦波成分に関するターゲット属性データ（平均アンプ静的成分Ａtar-sync-sta、平均アンプビブラート的成分Ａtar-sync-vib、ピッチ静的成分Ｐtar-sync-sta、ピッチビブラート的成分Ｐtar-sync-vib及びスペクトラル・シェイプＳtar-sync(f)）及びコントローラ２９から入力される正弦波成分属性データ選択情報に基づいて、新しい正弦波成分属性データである新規アンプ成分Ａnew、新規ピッチ成分Ｐnew及び新規スペクトラル・シェイプＳnew(f)を生成する。

【００８１】すなわち、新規アンプ成分Ａnewについては、次式により生成する。

Ａnew＝Ａ*-sta＋Ａ*-vib（ただし、*は、me又はtar-sync）

より具体的には、図８（Ｄ）に示すように、新規アンプ成分Ａnewを元属性データの平均アンプ静的成分Ａme-staあるいはターゲット属性データの平均アンプ静的成分Ａtar-sync-staのいずれか一方及び元属性データの平均アンプビブラート的成分Ａme-vibあるいはターゲット属性データの平均アンプビブラート的成分Ａtar-sync-vibのいずれか一方の組み合わせとして生成する。

【００８２】また、新規ピッチ成分Ｐnewについては、次式により生成する。

Ｐnew＝Ｐ*-sta＋Ｐ*-vib（ただし、*は、me又はtar-sync）

より具体的には、図８（Ｄ）に示すように、新規ピッチ成分Ｐnewを元属性データのピッチ静的成分Ｐme-staあるいはターゲット属性データのピッチ静的成分Ｐtar-sync-staのいずれか一方及び元属性データのピッチビブラート的成分Ｐme-vibあるいはターゲット属性データのピッチビブラート的成分Ｐtar-sync-vibのいずれか一方の組み合わせとして生成する。

【００８３】また、新規スペクトラル・シェイプＳnew(f)については、次式により生成する。

Ｓnew(f)＝Ｓ*(f)（ただし、*は、me又はtar-sync）

ところで、一般的にアンプ成分が大きい場合には、高域まで伸びた抜けの明るい音となり、アンプ成分が小さい場合には、逆にこもった音になる。そこで、新規スペクトラル・シェイプＳnew(f)に関しては、このような状態をシミュレートすべく、図１１に示すように、スペクトラル・シェイプの高域成分、すなわち、高域成分部分のスペクトラル・シェイプの傾きを新規アンプ成分Ａnewの大きさに応じて補償するスペクトラルチルト補償（spectral tilt correction）を行って、コントロールすることにより、よりリアルな音声を再生することができる。

【００８４】続いて、生成された新規アンプ成分Ａnew、新規ピッチ成分Ｐnew及び新規スペクトラル・シェイプＳnew(f)について、必要に応じてコントローラ２９から入力される正弦波成分属性データ変形情報に基づいて、属性データ変形部２４によりさらなる変形を行う。例えば、スペクトラル・シェイプを全体的に間延びさせる等の変形を行う。

【００８５】［３．１０］　残差成分選択部の動作

一方、残差成分選択部２５は、イージーシンクロナイゼーション処理部２２から入力された置換済ターゲットフレーム情報データＩＮＦtar-syncに含まれるターゲット属性データのうち残差成分に関するターゲット属性データ（残差成分Ｒtar-sync(f)）、残差成分保持部１２に保持されている残差成分信号（周波数波形）Ｒme(f)及びコントローラ２９から入力される残差成分属性データ選択情報に基づいて新しい残差成分属性データである新規残差成分Ｒnew(f)を生成する。

【００８６】すなわち、新規残差成分Ｒnew(f)については、次式により生成する。

Ｒnew(f)＝Ｒ*(f)（ただし、*は、me又はtar-sync）

この場合においては、me又はtar-syncのいずれを選択するかは、新規スペクトラル・シェイプＳnew(f)と同一のものを選択するのがより好ましい。

【００８７】さらに、新規残差成分Ｒnew(f)に関しても、新規スペクトラル・シェイプと同様な状態をシミュレートすべく、図１１に示したように、残差成分の高域成分、すなわち、高域成分部分の残差成分の傾きを新規アンプ成分Ａnewの大きさに応じて補償するスペクトラルチルト補償（spectral tilt correction）を行って、コントロールすることにより、よりリアルな音声を再生することができる。

【００８８】［３．１１］　正弦波成分生成部の動作

続いて、正弦波成分生成部２６は、属性データ変形部２４から出力された変形を伴わない、あるいは、変形を伴う新規アンプ成分Ａnew、新規ピッチ成分Ｐnew及び新規スペクトラル・シェイプＳnew(f)に基づいて、当該フレ

ームにおける新たな正弦波成分（Ｆ”０、Ａ”０）、（Ｆ”１、Ａ”１）、（Ｆ”２、Ａ”２）、……、（Ｆ”(N-1)、Ａ”(N-1)）のＮ個の正弦波成分（以下、これらをまとめてＦ”ｎ、Ａ”ｎと表記する。ｎ＝０～（Ｎ－１）。）を求める。

【００８９】より具体的には、次式により新規周波数Ｆ”ｎおよび新規アンプＡ”ｎを求める。

Ｆ”ｎ＝Ｆ’ｎ×Ｐnew

Ａ”ｎ＝Ｓnew(Ｆ”ｎ）×Ａnew

なお、完全倍音構造のモデルとして捉えるのであれば、

Ｆ”ｎ＝（ｎ＋１）×Ｐnew

となる。

【００９０】［３．１２］　正弦波成分変形部の動作

さらに、求めた新規周波数Ｆ”ｎおよび新規アンプＡ”ｎについて、必要に応じてコントローラ２９から入力される正弦波成分変形情報に基づいて、正弦波成分変形部２７によりさらなる変形を行う。例えば、偶数倍音成分の新規アンプＡ”ｎ（＝Ａ”０、Ａ”２、Ａ”４、……）だけを大きく（例えば、２倍する）等の変形を行う。これによって得られる変換音声にさらにバラエティーを持たせることが可能となる。

【００９１】［３．１３］　逆高速フーリエ変換部の動作

次に逆高速フーリエ変換部２８は、求めた新規周波数Ｆ”ｎおよび新規アンプＡ”ｎ（＝新規正弦波成分）並びに新規残差成分Ｒnew(f)をＦＦＴバッファに格納し、順次逆ＦＦＴを行い、さらに得られた時間軸信号を一部重複するようにオーバーラップ処理し、それらを加算する加算処理を行うことにより新しい有声音の時間軸信号である変換音声信号を生成する。

【００９２】このとき、コントローラ２９から入力される正弦波成分／残差成分バランス制御信号に基づいて、正弦波成分及び残差成分の混合比率を制御し、よりリアルな有声信号を得る。この場合において、一般的には、残差成分の混合比率を大きくするとざらついた声が得られる。この場合において、ＦＦＴバッファに新規周波数Ｆ”ｎおよび新規アンプＡ”ｎ（＝新規正弦波成分）並びに新規残差成分Ｒnew(f)を格納するに際し、異なるピッチ、かつ、適当なピッチで変換された正弦波成分をさらに加えることにより変換音声信号としてハーモニーを得ることができる。さらにシーケンサ３１により伴奏音に適合したハーモニーピッチを与えることにより、伴奏に適合した音楽的ハーモニーを得ることができる。

【００９３】［３．１４］　クロスフェータの動作

次にクロスフェータ部３０は、元無声／有声検出信号Ｕ／Ｖme(t)に基づいて、入力音声信号Ｓvが無声（Ｕ）である場合には、入力音声信号Ｓvをそのままミキサ３３に出力する。また、入力音声信号Ｓvが有声（Ｖ）である場合には、逆高速フーリエ変換部２８が出力した変換音声信号をミキサ３３に出力する。この場合において、切替スイッチとしてクロスフェータ部３０を用いているのは、クロスフェード動作を行わせることによりスイッチ切替時のクリック音の発生を防止するためである。

【００９４】［３．１５］　シーケンサ、音源部、ミキサ及び出力部の動作

一方、シーケンサ３１は、カラオケの伴奏音を発生するための音源制御情報を例えば、ＭＩＤＩ（Musical Instrument Digital Interface）データなどとして音源部３２に出力する。これにより音源部３２は、音源制御情報に基づいて伴奏信号を生成し、ミキサ３３に出力する。ミキサ３３は、入力音声信号Ｓvあるいは変換音声信号のいずれか一方及び伴奏信号を混合し、混合信号を出力部３４に出力する。出力部３４は、図示しない増幅器を有し混合信号を増幅して音響信号として出力することとなる。

【００９５】［４］　実施形態の変形例

［４．１］　第１変形例

以上の説明においては、属性データとしては、元属性データあるいはターゲット属性データのいずれかを選択的に用いる構成としていたが、元属性データ及びターゲット属性データの双方を用い、補間処理を行うことにより中間的な属性を有する変換音声信号を得るように構成することも可能である。しかしながら、このような構成によれば、ものまねをしようとする歌唱者及びものまねの対象（target）となる歌唱者のいずれにも似ていない変換音声が得られる場合もある。また、特にスペクトラル・シェイプを補間処理によって求めた場合には、ものまねをしようとする歌唱者が「あ」を発音し、ものまねの対象となる歌唱者が「い」を発音している場合などには、「あ」でも「い」でもない音が変換音声として出力される可能性があり、その取扱には注意が必要である。

【００９６】［４．２］　第２変形例

正弦波成分の抽出は、この実施形態で用いた方法に限らない。要は、音声信号に含まれる正弦波を抽出できればよい。

【００９７】［４．３］　第３変形例

本実施形態においては、ターゲットの正弦波成分及び残差成分を記憶したが、これに換えて、ターゲットの音声そのものを記憶し、それを読み出してリアルタイム処理によって正弦波成分と残差成分とを抽出してもよい。すなわち、本実施形態でものまねをしようとする歌唱者の音声に対して行った処理と同様の処理をターゲットの歌唱者の音声に対して行ってもよい。

【００９８】［４．４］　第４変形例

本実施形態においては、属性データとして、ピッチ、アンプ、スペクトラル・シェイプの全てを取り扱ったが、少なくともいずれか一つを扱うようにすることも可能である。

【００９９】［４．５］　第５変形例

本実施形態の補間合成部９におけるデータ構成は、その

他の各部（例えばピッチ正規化部１７～正弦波成分変形部２７に至る区間）におけるデータ構成としても良いことは言うまでもない。特に、ターゲットフレーム情報保持部２０においては、１曲分のターゲットフレーム情報が記憶されるため、上記データ構成を用いることによるデータ量の削減効果が大きい。

【０１００】［５］　実施形態の効果

以上のように、本実施形態によれば、周波数Ｆｋに対して差分値ｄＦｋ、比率ｒＦｋまたは対数値ｃＦｋを記憶するようにしたため、僅かな記憶容量で高精度な音声特徴情報を記憶できる。さらに、アンプＡｋに対して１バイト中の７ビットを割り当てて１２８ｄＢのダイナミックレンジを確保するとともに残りの１ビットにおいて周波数を特定するための情報（上記対数値ｃＦｋ等）が存在するか否かを示す態様においては、さらに記憶容量を削減することができる。また、アンプＡｋに代えて「ｎＡｋ＝α・（Ａｋ＋β）」によるアンプｎＡｋを記憶することにより、所望のダイナミックレンジに応じて可能な限り高い分解能を確保することが可能である。

【０１０１】

【発明の効果】以上説明したように本発明によれば、各乗算結果Ｆ０×ｋと各周波数Ｆｋとの差分または割合を記憶することによって音声の特徴を記憶するから、僅かな記憶容量で高精度な音声特徴情報を記憶できる。

【図面の簡単な説明】

【図１】　本発明の一実施形態の構成を示すブロック図（その１）である。

【図２】　本発明の一実施形態の構成を示すブロック図（その２）である。

【図３】　実施形態におけるフレームの状態を示す図である。

【図４】　実施形態における周波数スペクトルのピーク＊検出を説明するための説明図である。

【図５】　実施形態におけるフレーム毎のピーク値の連携を示す図である。

【図６】　実施形態における周波数値の変化状態を示す図である。

【図７】　実施形態における処理過程における確定成分の変化状態を示す図である。

【図８】　実施形態における信号処理の説明図である。

【図９】　イージーシンクロナイゼーション処理のタイミングチャートである。

【図１０】　イージーシンクロナイゼーション処理フローチャートである。

【図１１】　スペクトラル・シェイプのスペクトラルチルト補償について説明する図である。

【符号の説明】

１…マイク、２…分析窓生成部、３…入力音声信号切出部、４…高速フーリエ変換部、５…ピーク検出部、６…無声／有声検出部、７…ピッチ抽出部、８…ピーク連携部、９…補間合成部、１０…残差成分検出部、１１…高速フーリエ変換部、１２…残差成分保持部、１３…正弦波成分保持部、１４…平均アンプ演算部、１５…アンプ正規化部、１６…スペクトラル・シェイプ演算部、１７…ピッチ正規化部、１８…元フレーム情報保持部、１９…静的変化／ビブラート的変化分離部、２０…ターゲットフレーム情報保持部、２１…キーコントロール／テンポチェンジ部、２２…イージーシンクロナイゼーション処理部、２３…正弦波成分属性データ選択部、２４…属性データ変形部、２５…残差成分選択部、２６…正弦波成分生成部、２７…正弦波成分変形部、２８…逆高速フーリエ変換部、２９…コントローラ、３０…クロスフェーダ部、３１…シーケンサ、３２…音源部、３３…ミキサ、３４…出力部。

Fig.3　【図３】

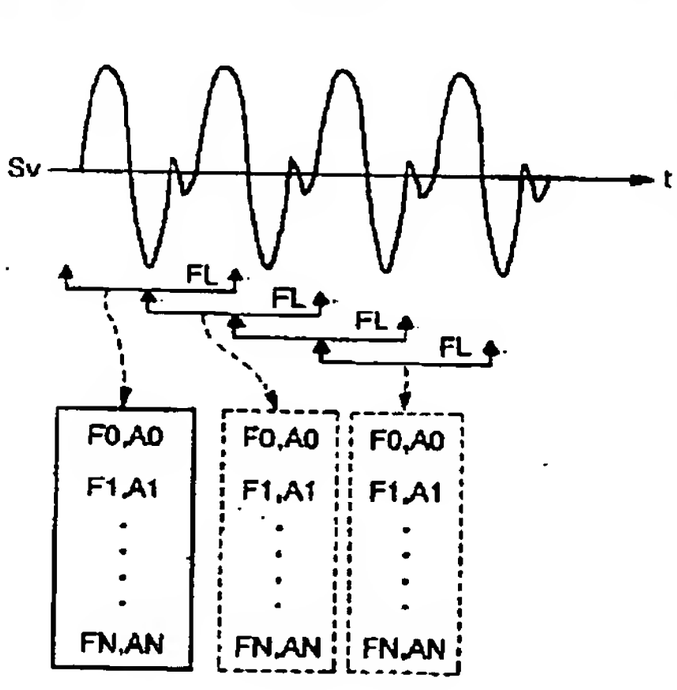


Fig.4　【図４】

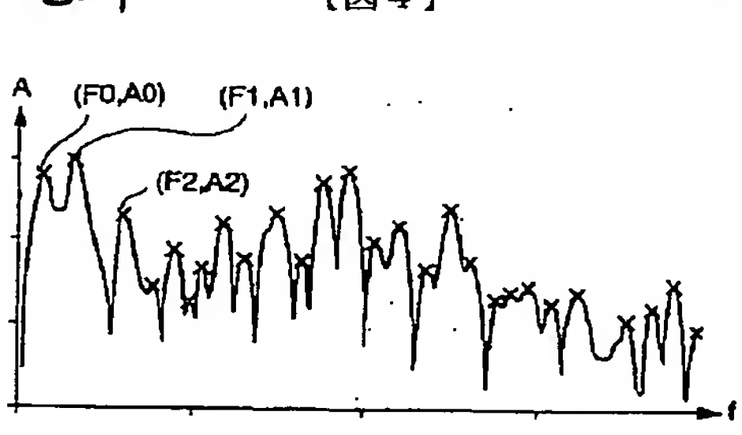


Fig.11　【図１１】

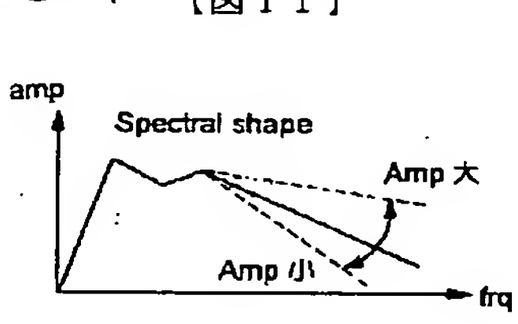
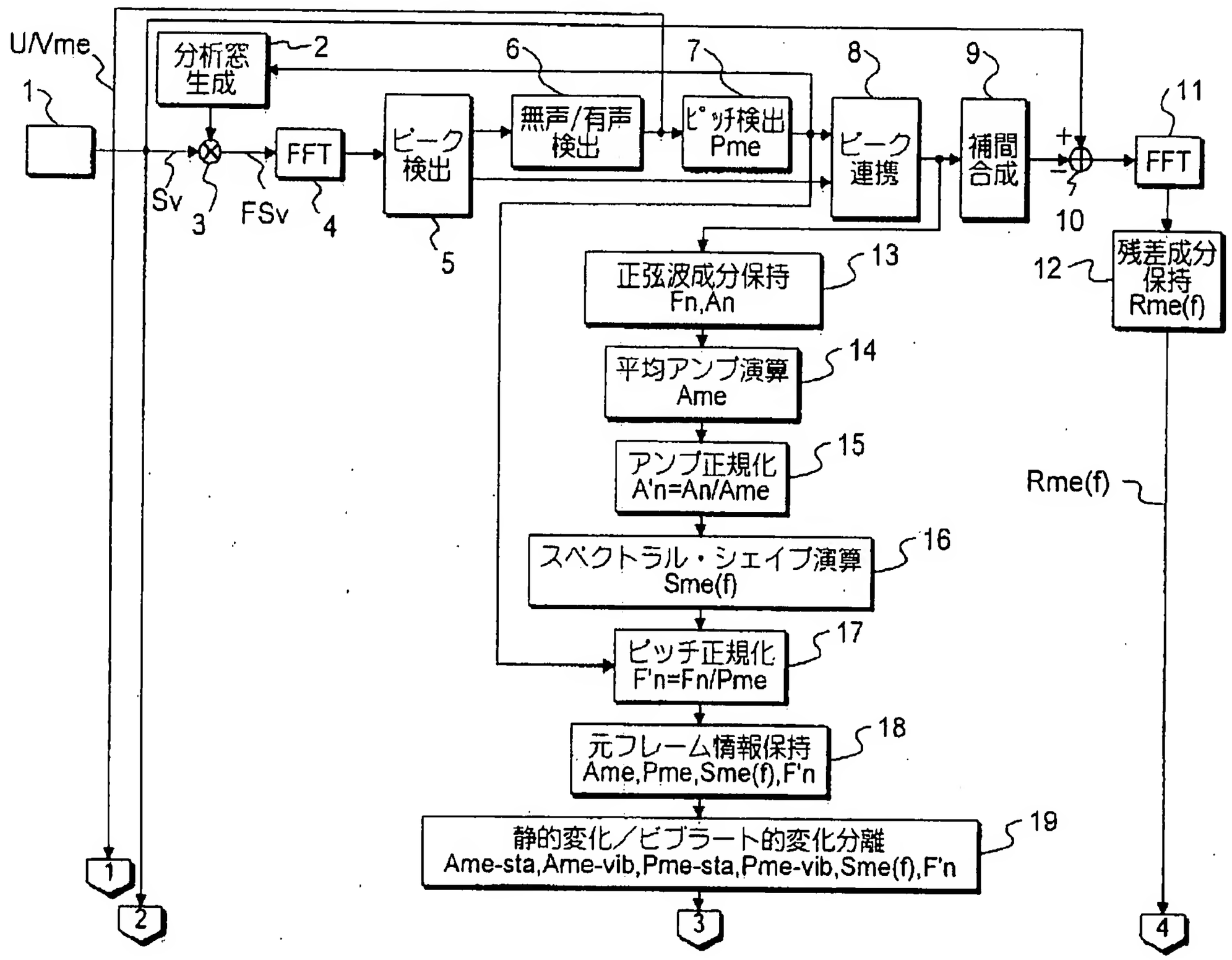


Fig. /


U/Vme
1
分析窓生成
2
Sv
3
FSv
FFT
4
ピーク検出
5
無声/有声検出
6
ピッチ検出 Pme
7
ピーク連携
8
補間合成
9
+
−
10
FFT
11
残差成分保持 Rme(f)
12
Rme(f)
正弦波成分保持 Fn,An
13
平均アンプ演算 Ame
14
アンプ正規化 A'n=An/Ame
15
スペクトラル・シェイプ演算 Sme(f)
16
ピッチ正規化 F'n=Fn/Pme
17
元フレーム情報保持 Ame,Pme,Sme(f),F'n
18
静的変化／ビブラート的変化分離 Ame-sta,Ame-vib,Pme-sta,Pme-vib,Sme(f),F'n
19
1
2
3
4

【図１】

Fig. 2

【図２】

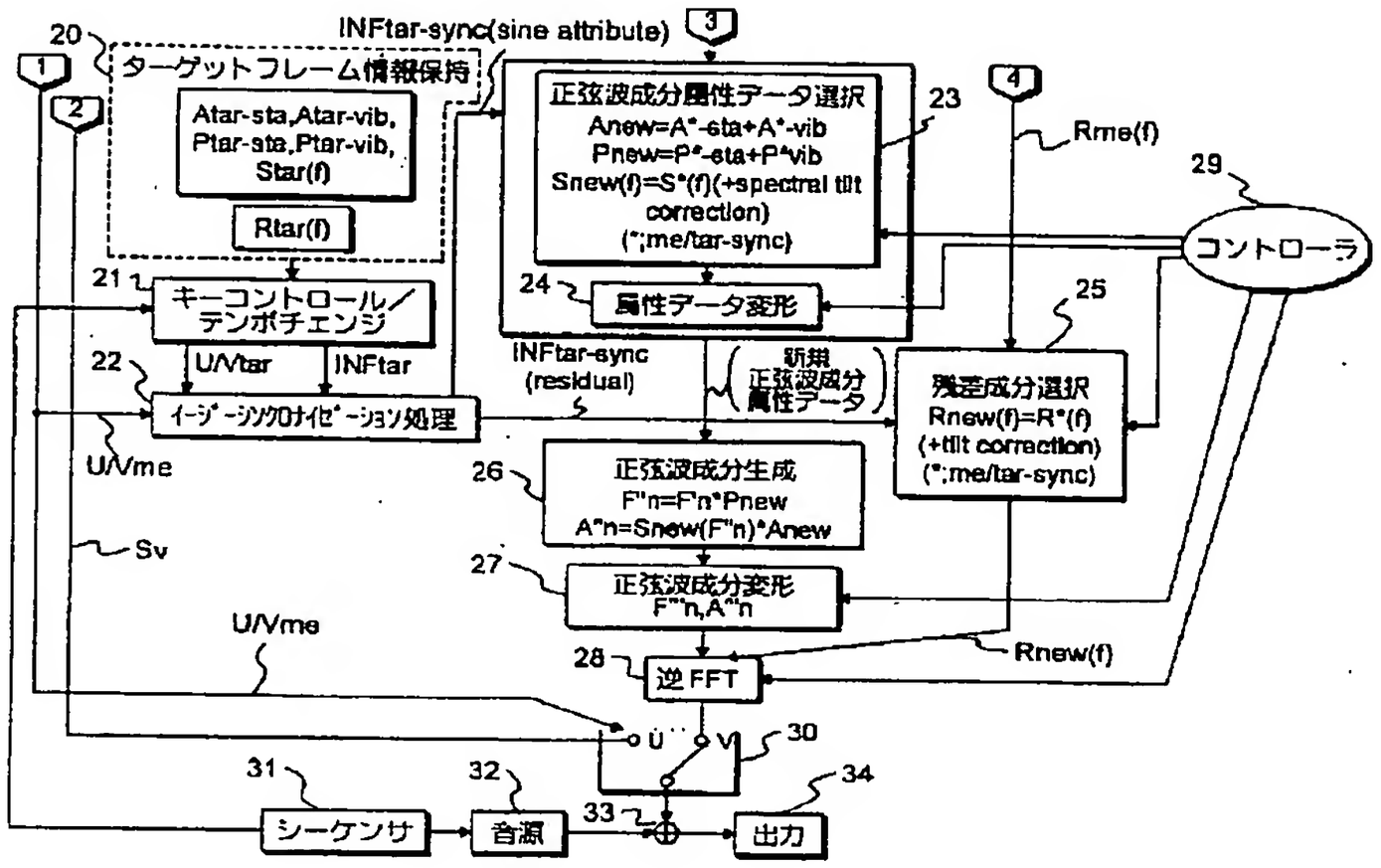


Fig. 5

【図５】

Fig. 6

【図６】

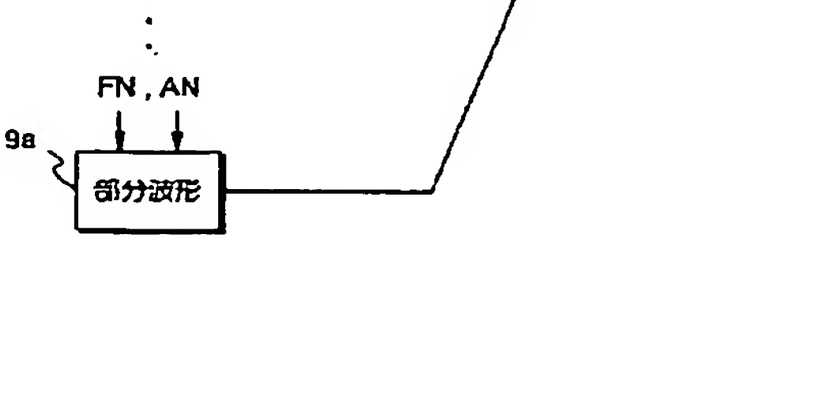


Fig. 7

【図７】

Fig. 8

【図８】

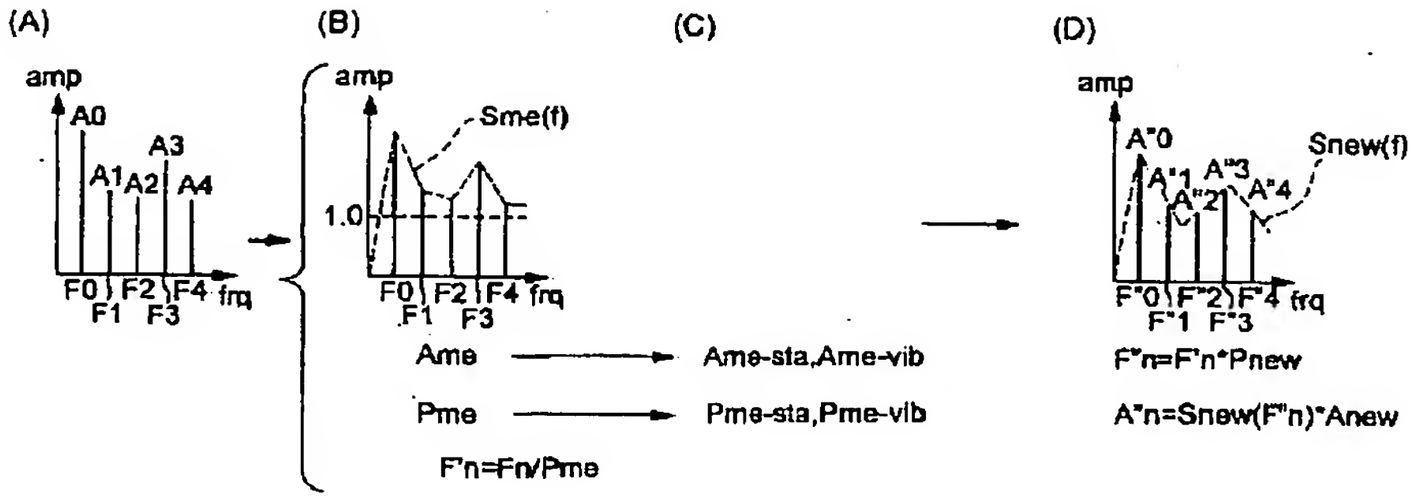


Fig. 9

【図９】

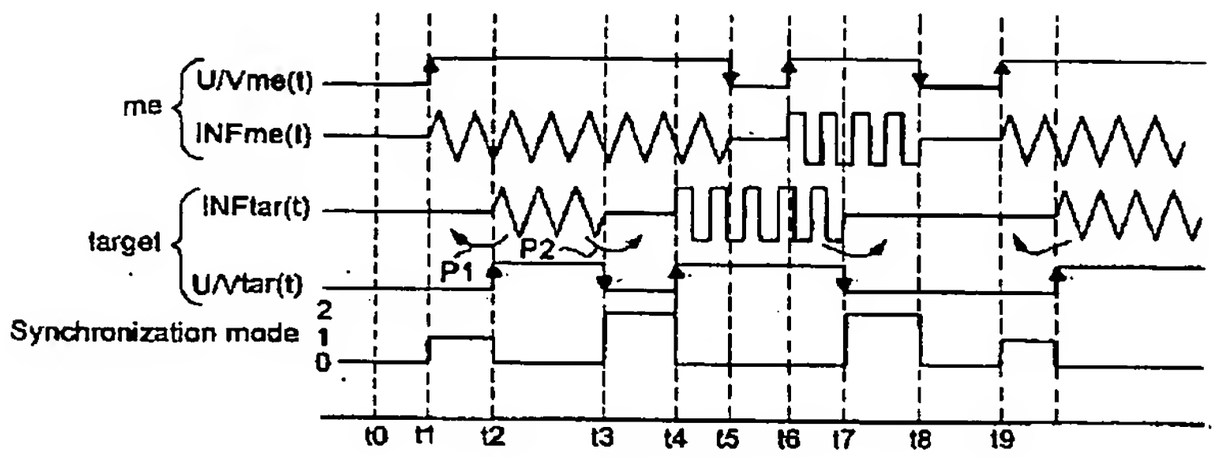


Fig. 10

【図１０】